Data Analyst Project

By Babita Mehra

Introduction

Ola is one of India's largest ride-sharing and transportation platforms, offering services to millions of customers across the country.

I have performed a comprehensive analysis of Ola's operational data to extract valuable insights related to revenue generation, customer ratings, ride cancellations, and vehicle types.

This project utilizes ***SQL*** to handle and analyse raw datasets and ***Power BI*** for data visualisation and storytelling. The goal is to assist Ola's business team in understanding key performance area and identifying potential improvements.

Key Objectives

- To analyze total revenue generated from rides.
- To identify ride cancellation patterns(both by drivers & customers) based on reasons.
- To explore customer ratings and determine which factors might influence them.
- To study the performance of different vehicle types(Prime Sedan, Prime Plus, Prime SUV, Auto, Bike, e-bike, Mini).
- To create interactive and visual dashboards using Power BI for better understanding and reporting.

Tooles Used

1. **Microsoft Excel**: Used for data cleaning, pre-processing, and transforming raw data into more structured format before querying.
2. **SQL**: For data cleaning, querying, and aggregating large datasets.
3. **Power BI**: For building dashboards and visualizing key findings.

OLA Data Analyst Project

SQL Questions:

1. Retrieve all successful bookings:
2. Find the average ride distance for each vehicle type:
3. Get the total number of cancelled rides by customers:
4. List the top 5 customers who booked the highest number of rides:
5. Get the number of rides cancelled by drivers due to personal and car-related issues:
6. Find the maximum and minimum driver ratings for Prime Sedan bookings:
7. Retrieve all rides where payment was made using UPI:
8. Find the average customer rating per vehicle type:
9. Calculate the total booking value of rides completed successfully:
10. List all incomplete rides along with the reason:

Power BI Questions:

1. Ride Volume Over Time
2. Booking Status Breakdown
3. Top 5 Vehicle Types by Ride Distance
4. Average Customer Ratings by Vehicle Type
5. cancelled Rides Reasons
6. Revenue by Payment Method
7. Top 5 Customers by Total Booking Value
8. Ride Distance Distribution Per Day
9. Driver Ratings Distribution
10. Customer vs. Driver Ratings

Data Columns

1. Date
2. Time
3. Booking_ID
4. Booking_Status
5. Customer_ID
6. Vehicle_Type
7. Pickup_Location
8. Drop_Location
9. V_TAT
10. C_TAT
11. cancelled_Rides_by_Customer
12. cancelled_Rides_by_Driver
13. Incomplete_Rides
14. Incomplete_Rides_Reason
15. Booking_Value
16. Payment_Method
17. Ride_Distance
18. Driver_Ratings
19. Customer_Rating

OLA Data Analyst Project

SQL Answers:

1. Retrieve all successful bookings:

```
SELECT * FROM bookings WHERE Booking_Status = 'Success';
```

2. Find the average ride distance for each vehicle type:

```
SELECT Vehicle_Type, AVG(Ride_Distance) as avg_distance FROM bookings GROUP BY Vehicle_Type;
```

3. Get the total number of cancelled rides by customers:

```
SELECT COUNT(*) FROM bookings WHERE Booking_Status = 'cancelled by Customer';
```

4. List the top 5 customers who booked the highest number of rides:

```
SELECT Customer_ID, COUNT(Booking_ID) as total_rides FROM bookings GROUP BY Customer_ID ORDER BY total_rides DESC LIMIT 5;
```

5. Get the number of rides cancelled by drivers due to personal and car-related issues:

```
SELECT COUNT(*) FROM bookings WHERE cancelled_Rides_by_Driver = 'Personal & Car related issue';
```

6. Find the maximum and minimum driver ratings for Prime Sedan bookings:

```
SELECT MAX(Driver_Ratings) as max_rating, MIN(Driver_Ratings) as min_rating FROM bookings WHERE Vehicle_Type = 'Prime Sedan';
```

7. Retrieve all rides where payment was made using UPI:

```
SELECT * FROMbookings WHERE Payment_Method = 'UPI';
```

8. Find the average customer rating per vehicle type:

```
SELECT Vehicle_Type, AVG(Customer_Rating) as avg_customer_rating FROM bookings GROUPBYVehicle_Type;
```

9. Calculate the total booking value of rides completed successfully:

```
SELECT SUM(Booking_Value) as total_successful_value FROM bookings WHERE Booking_Status = 'Success';
```

10. List all incomplete rides along with the reason:

```
SELECT Booking_ID, Incomplete_Rides_Reason FROM bookings WHERE Incomplete_Rides = 'Yes';
```

Power BI Answers:

Segregation of the views:

1. Overall–

Ride Volume Over Time

Booking Status Breakdown

2. Vehicle Type –

Top 5 Vehicle Types by Ride Distance

3. Revenue –

Revenue by Payment Method

Top 5 Customers by Total Booking Value

Ride Distance Distribution Per Day

4. Cancellation—

Cancelled Rides Reasons (Customer)

cancelled Rides Reasons(Drivers)

5. Ratings—

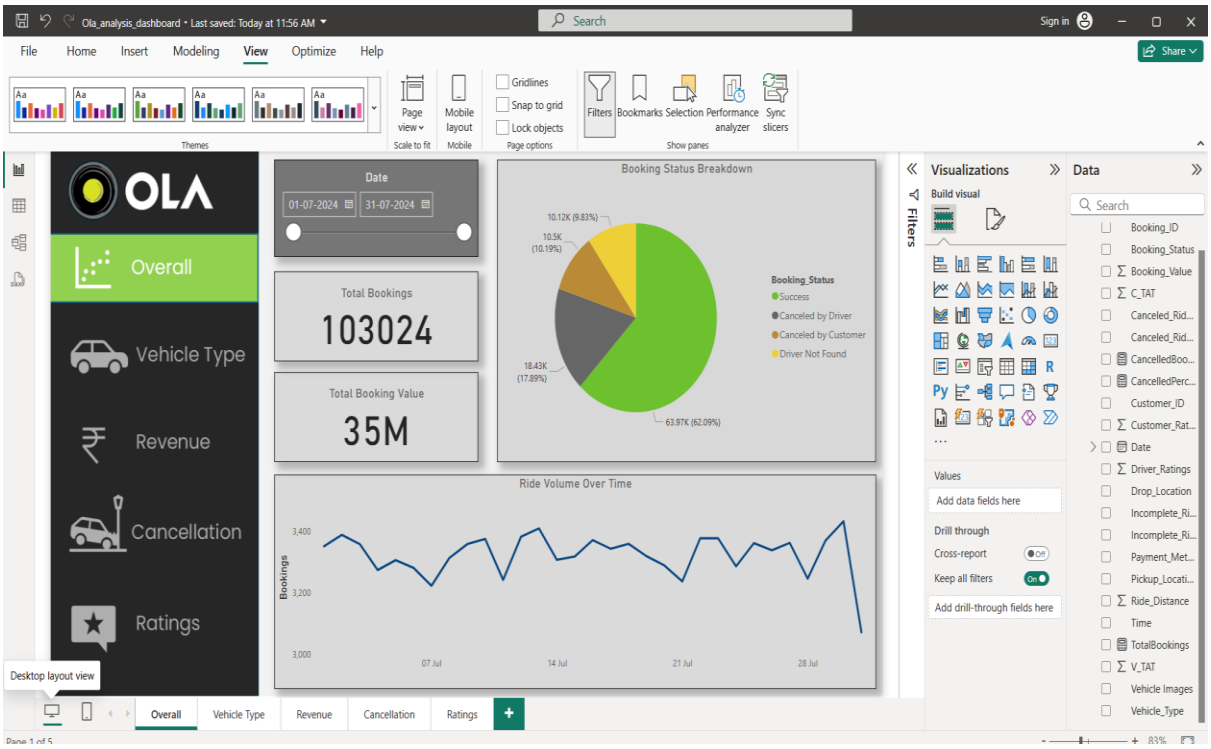
Driver Ratings

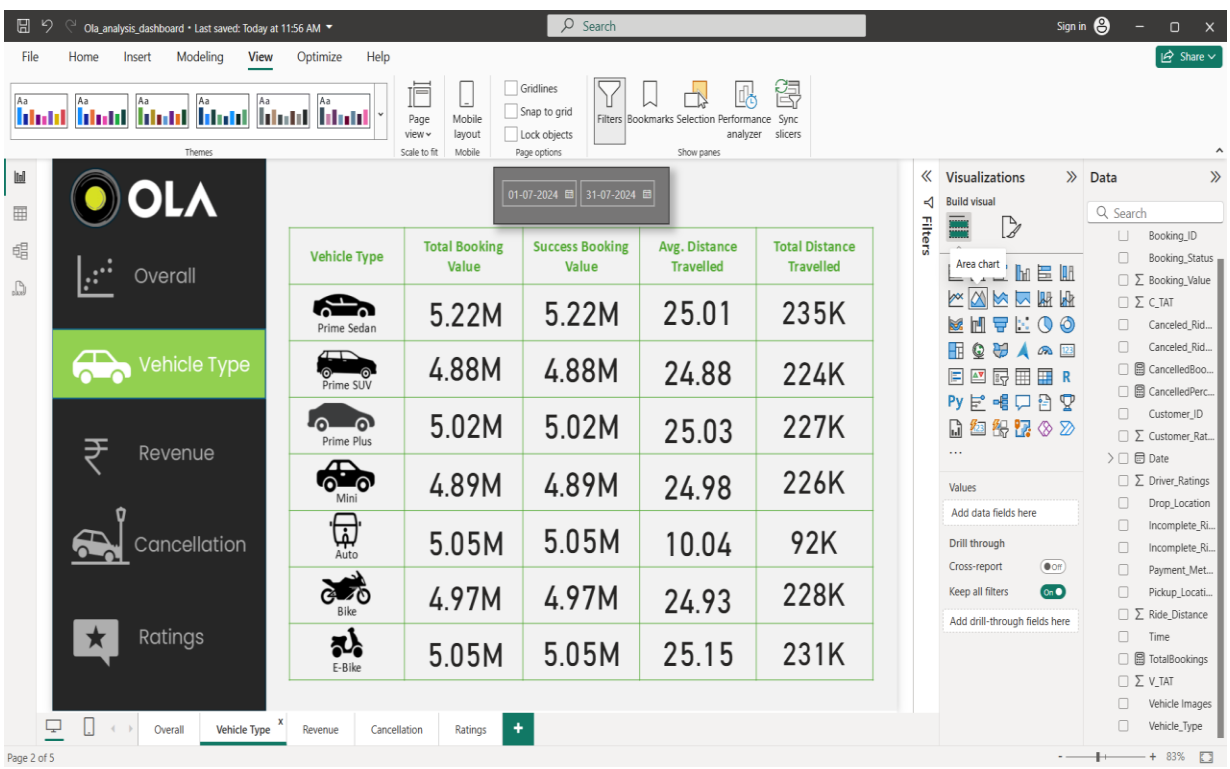
Customer Ratings

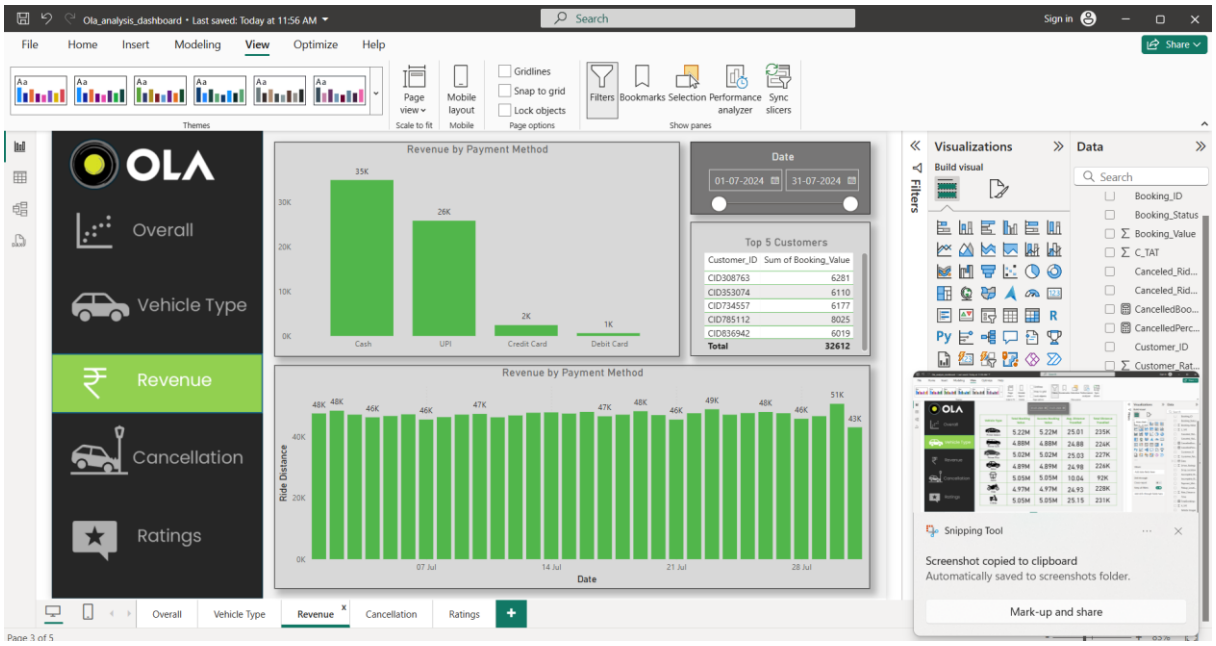
Answers:

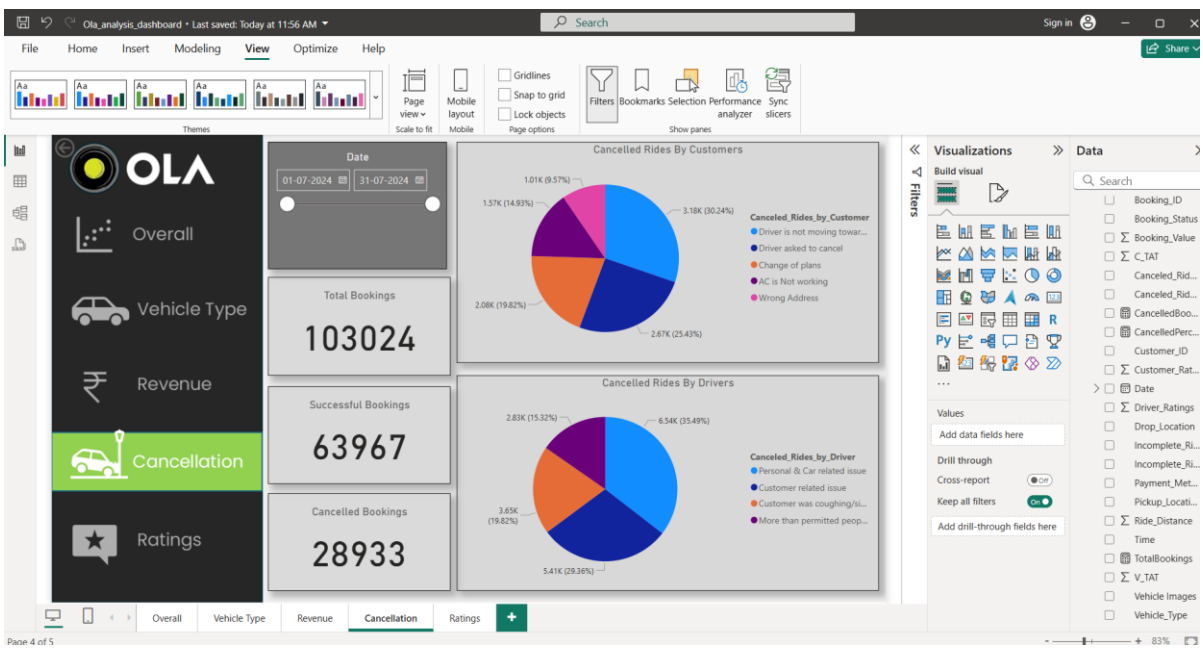
1. Ride Volume Over Time: A time-series chart showing the number of rides per day/week.
2. Booking Status Breakdown: A pie or doughnut chart displaying the proportion of different booking statuses (success, cancelled by the customer, cancelled by the driver, etc.).
3. Top 5 Vehicle Types by Ride Distance: A bar chart ranking vehicle types based on the total distance covered.
4. Average Customer Ratings by Vehicle Type: A column chart showing the average customer ratings for different vehicle types.
5. cancelled Rides Reasons: A bar chart that highlights the common reasons for ride cancellations by customers and drivers.
6. Revenue by Payment Method: A stacked bar chart displaying total revenue based on payment methods (Cash, UPI, Credit Card, etc.).
7. Top 5 Customers by Total Booking Value: A leaderboard visual listing customers who have spent the most on bookings.
8. Ride Distance Distribution Per Day: A histogram or scatter plot showing the distribution of ride distances for different Dates.
9. Driver Rating Distribution: A box plot visualizing the spread of driver ratings for different vehicle types.
10. Customer vs. Driver Ratings: A scatter plot comparing customer and driver ratings for each completed ride, analyzing correlations.

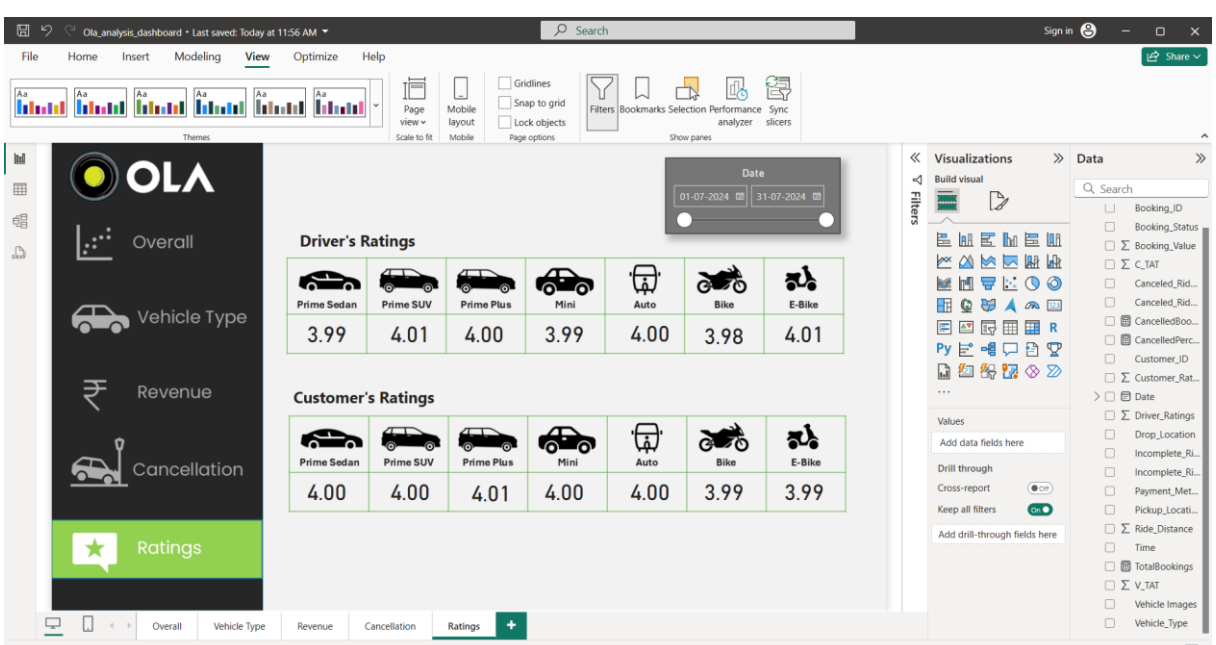
Dasboard : Power BI






OLA
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Revenue by Payment Method
35K
26K
2K
1K
Cash
UPI
Credit Card
Debit Card
Date
01-07-2024
31-07-2024
Top 5 Customers
Customer_ID Sum of Booking_Value
CID308763 6281
CID353074 6110
CID734557 6177
CID785112 8025
CID836942 6019
Total 32612
Revenue by Payment Method
Ride Distance
Date
07 Jul
14 Jul
21 Jul
28 Jul
Visualizations
Build visual
Data
Search
Booking_ID
Booking_Status
Booking_Value
C_TAT
Canceled_Rid...
Canceled_Rid...
CancelledBoo...
CancelledPerc...
Customer_ID
Customer_Rat...
Snipping Tool
Screenshot copied to clipboard
Automatically saved to screenshots folder.
Mark-up and share
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Page 3 of 5


OLA
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Date
01-07-2024
31-07-2024
Total Bookings
103024
Successful Bookings
63967
Cancelled Bookings
28933
Cancelled Rides By Customers
1.01K (9.57%)
1.57K (14.93%)
3.18K (30.24%)
2.08K (19.82%)
2.67K (25.43%)
Canceled_Rides_by_Customer
Driver is not moving towar...
Driver asked to cancel
Change of plans
AC is Not working
Wrong Address
Cancelled Rides By Drivers
2.83K (15.32%)
6.54K (35.49%)
3.65K (19.82%)
5.41K (29.36%)
Canceled_Rides_by_Driver
Personal & Car related issue
Customer related issue
Customer was coughing/si...
More than permitted peop...
Visualizations
Build visual
Data
Search
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Page 4 of 5


OLA
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Date
01-07-2024
31-07-2024
Driver's Ratings
Prime Sedan 3.99
Prime SUV 4.01
Prime Plus 4.00
Mini 3.99
Auto 4.00
Bike 3.98
E-Bike 4.01
Customer's Ratings
Prime Sedan 4.00
Prime SUV 4.00
Prime Plus 4.01
Mini 4.00
Auto 4.00
Bike 3.99
E-Bike 3.99
Visualizations
Build visual
Data
Search
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Overall
Vehicle Type
Revenue
Cancellation
Ratings
Page 5 of 5

conclusion

Through this project on OLA's ride data, we successfully analyzed key business metrics using SQL aggregate functions and data visualization through dashboards. The analysis revealed important patterns:

- High Revenue & Ride Demand: Certain cities consistently show higher ride demand and revenue, indicating strong market presence in these areas.
- Driver & Customer Behavior: We identified trends in driver activity and customer preferences, such as peak booking hours and frequent ride types.
- Performance Insights: Using functions like SUM(), AVG(), COUNT(), and applying GROUP BY and HAVING, we uncovered insights into average trip fare, total bookings, and the distribution of rides.
- Operational Opportunities: The dashboard highlights regions where service performance can be improved and areas with growth potential.

In summary, the project provided valuable business intelligence for OLA, helping to make data-driven decisions that can improve customer satisfaction, optimize driver allocation, and enhance revenue streams.

This project demonstrates how SQL and data visualization empower real-world business solutions.